来沪学术报告之八

计 算 机 图 象 识 别

（内部资料·注意保存）

上海科学技术情报研究所

一九七三年九月

说　　明

美国加利福尼亚州美籍中国科学家访问团团员王佑曾，一九五五年毕业于美国普林斯理大学电机系。曾在国际商业机器公司任职，以后历任加利福尼亚大学柏来克分校电机系副教授、教授，现任电机系副主任。

一九七三年七月，王佑曾在访问期间，曾在本市作了关于计算机图象识别的学术报告。这次报告会由华东计算技术研究所瞿兆荣同志主持。会议记录由上海电机综合研究所、上海计算技术研究所负责整理，但未经报告人审阅。此份资料仅供参考，请勿翻印。

上海科学技术情报研究所

计 算 机 图 象 识 别

有人把计算机称为电脑，这是不对的。现在美国做机器人的，第一家是麻省理工学院。还有两家也做。有许多人在作定理证明，还有一种是用在下棋，打牌每一步有不少走法，用数学计算不行，要用更聪明的办法。最近美国有个下棋冠军写了一个程序，已经能和人下了，但只是开局和中局能下，到后来就乱了。

图象识别，很早就有人感兴趣了，我在1960年在IBM公司就研究了，图象识别是个很大很广的问题，应该分一分。

(1) 文字识别：这是做的最长久，最实用，也是最成功的一种，最大的应用是计算机的输入问题。

(2) 语言识别：如果成功的话也是很有用的，解决这个问题需要语言学相应的发展。当前的成绩还差，最大的困难是一个一个字是不能分的，现还未解决。原来想用语音来识别。很多用的是一个字一个字地讲，这只能用于小机器，大机器一边讲一边输出就要几个小时。

(3) 波形识别：比上两种容易，但它不容易找出一个模型。医学上的脑电图，心电图很有希望用计算机来识别。这方面开始作的相当好，最近没有消息。因为通常前80%容易，后20%就困难了。

(4) 图象识别：用于象癌切片处理，可以由机器图象识别来作。在技术上有几种方法：

1) 统计学的：这是最早的，用这种方法要有统计模型。一般用一些简单的如距离。已经作成的大部分是简单的模型。

2) 学习的：用“学习”这词评价太高了。近期一个重要的工作是Cover做的。

3) 语言的。

4) 拓扑学的：

我自己在感知器方面做了一些。

对于数学，历史上很早就有“投针问题”。有二组等距平行线，一组间距 α，一组间距 β，$\alpha<\beta$，不知道它们之间的转角，怎么来识别他们？可以分别用针在两组平行线上投，记下针和平行线相交的次数之比，做的次数充分多每次求和各收敛到一极限值，比较这两个极限值就可以识别那一组是 α 那一组是 β 了。

一般地图形可用随意地直线去划，相交的长求平均

$$\frac{1}{N}\sum_{n=1}^{N}\ln \xrightarrow[N\to\infty]{} t\mathrm{Orea}(C)$$

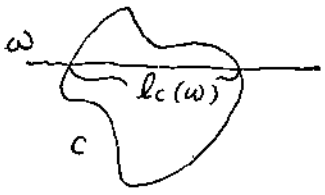
原理与上同。

有人用 $\frac{1}{N}\sum_{n=1}^{N}\ln\Rightarrow M_1$，$\frac{1}{N}\sum_{n=1}^{n}\ln\Rightarrow M_2$ 来作，结果是失望的，究其原因，一是收敛慢，二是这两个极限不能很好地分辨。他们就把注意力放在识别上，其次才考虑不变性。下面介绍我的一个工作，分两方面：一是象位方法，二是特征抽取。

先讲象位方法：象位和不变性是有关的。

（下图为保持原意用原文）

Shape Recognition



Generate “random” lines ω_1, ω_2, ……

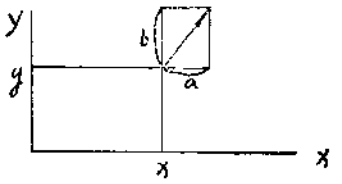
Measure $l_c(\omega_i)$, $i=1, 2,$ ……

Sample mean $\frac{1}{N}\sum_{n=1}^{N} l_c(\omega_n) \xrightarrow[N\to\infty]{} K(\mathrm{)rea}(C)$

Rigid body motions

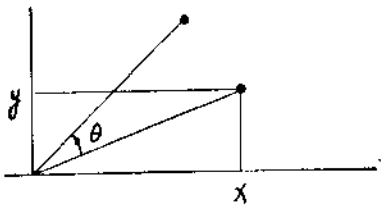
Jranslations: $t(a, b)$; $R^2\to R^2$

$$t(a, b)(x, y)=(x+a, y+b)$$



Rotation: τ_θ; $R^2\to R^2$

$$\tau_\theta\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$



G = Group of rigid body motions

$g \longleftrightarrow \tau_\theta t(a,\ b)$, $\theta \in [0,\ 2\pi)$ $(a,\ b) \in R^2$

$G \longleftrightarrow R^2 \times [0,\ 2\pi)$

$\mu(dg) = d\theta da db$ Invariant measure

Feature Space

Generator: ω_0

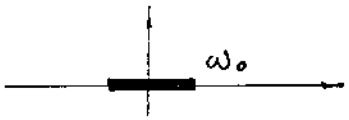
$\Omega = \{g\omega_0;\ g \in G\}$

$G_0 = \{g \in G;\ g\omega_0 = \omega_0\}$

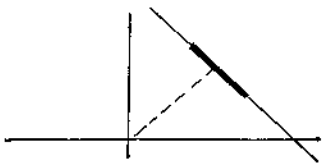
$\Omega \longleftrightarrow G/G_0$

$G \longleftrightarrow \Omega \times G_0$

Example



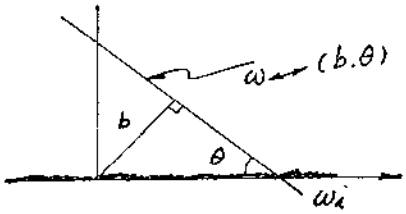
$G_0 = \{\tau_\pi, I\}$



Invariant measure

$$\mu_\Omega(d\omega) = \int_{G_0} f(g_0)\,\mu_G(d\omega \times dg_0)$$

Example



$\omega_0 = \{(x,\ y),\ y = 0,\ -\infty < x < +\infty\}$

$G_0 = \{\tau_\theta t(a,\ b);\ \theta - 0,\ \pi;$

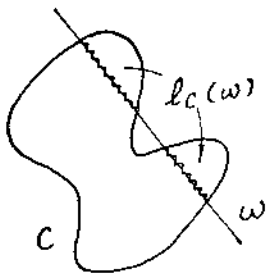
$-\infty < a < \infty,\ b = 0\}$

$\Omega \longleftrightarrow [0,\ \pi) \times (-\infty,\ \infty)$

$\mu_\Omega = db d\theta$

μ_Ω Can be normalized by introducing a retina $\tilde{\mu}_\Omega$ prob measure.

"Random" lines: lines generated according to $\tilde{\mu}_\Omega$ measure



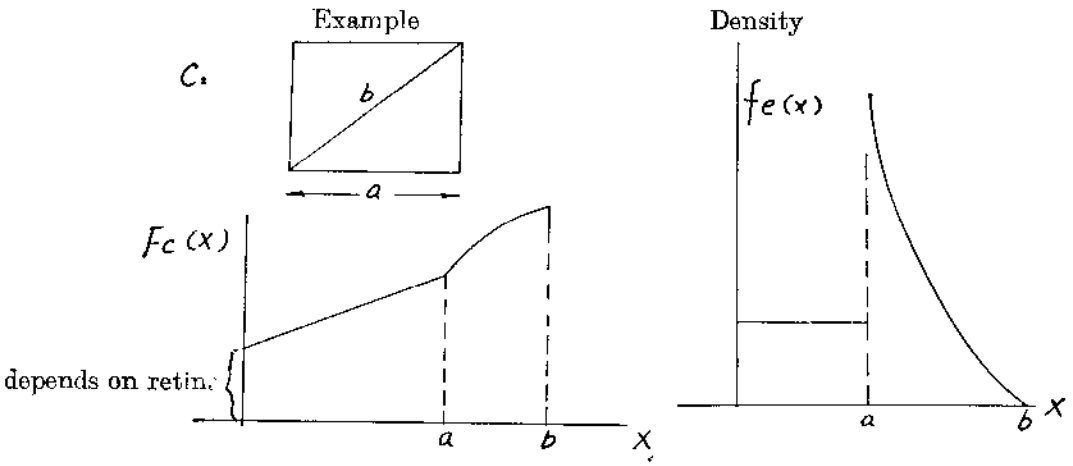
Chord length: $l_c(\omega)$

$$l_c(\omega) = l_{gc}(g\omega)$$

For a fixed C, $l_c(\omega)$, $\omega \in \Omega$; is a random variable under $\tilde{\mu}_\Omega$.

Distribution function $F_c(x) = \tilde{\mu}_\Omega(\{\omega, l(C, \omega) < x\})$

$$F_c = F_g(C)$$



Recognition Procedure

Generate Independent lines according to $\tilde{\mu}_\Omega$ probability measure.

$$\omega_1, \omega_2, \cdots\cdots$$

Observe: $l_c(\omega_1)$, $l_c(\omega_2)$, ……

indentically distributed independent random variables.

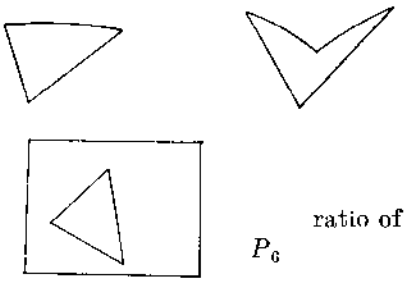
Decide: C

Standard statistical problem.

pairwisw discrimination

Example:

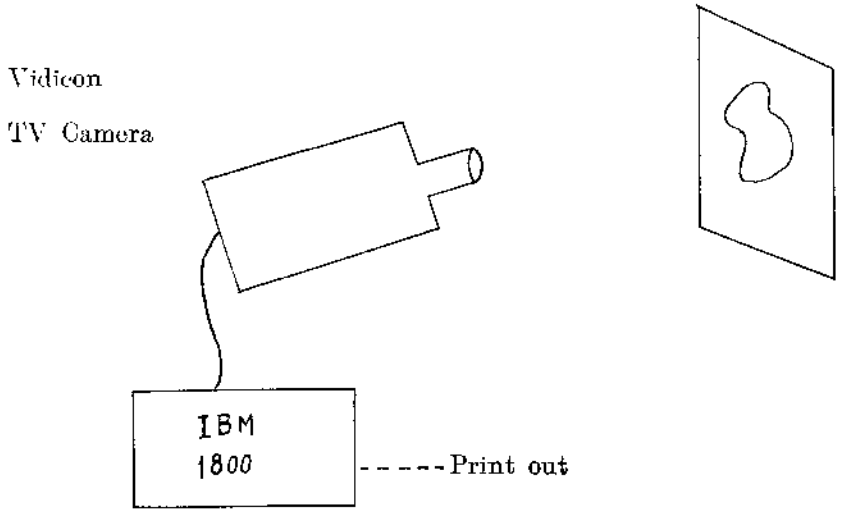
Sequential Likelihood Ratio Jest



error rate	aoerage sample size
0.01	~55
0.002	~70

在三年前我们算过一下，用十五个图象，有 105 个对，可以算出误差，决定了误差就可知平均子样，计算结果和 Boly 比较从四万条线变到三条线。当然可以不用直线，但实验表明直线最好，因为直线是局部的，而我们一般的图象看时也是局部的，直线还有个好处即容易作。我和一个学生作过实验，用一个旧的电视照相机改了一下，照相机容易做到大小不变，我们用的是很快地左右扫描。本来想在照相机里放一个照相机，但没有做，用了一个 IBM 1800。我们这个机器是可以转动的。

Experimental System



(1) Vidicon modified for vandom scanning

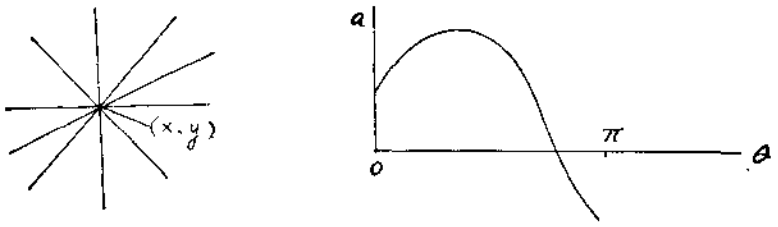
(2) Size Normalized by preliminavy raster scan.

下面讲特征抽取：

Representation of a point in feature space.

$$(x,\ y)\longleftrightarrow\{\omega\in\Omega:\ \omega\ni(x,\ y)\}$$

$$\longleftrightarrow\{(a,\ \theta):\ x\sin\theta+y\cos\theta=a\}$$

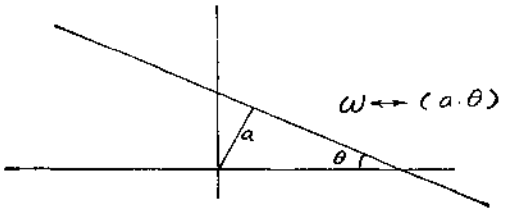


$$I(x,\ y)\ (a,\ \theta)=\begin{cases}1 & x\sin\theta+y\cos\theta=a\\ 0 & \text{other wise}\end{cases}$$

Feature Extraction

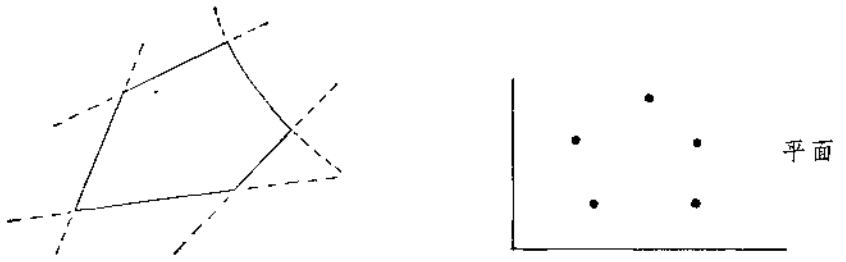
Example Feature = Straight line

Feature Space $\Omega\longleftrightarrow(-\infty,\ \infty)\times[0,\ \pi)$

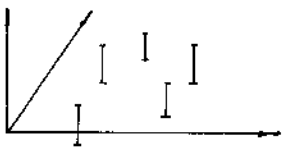


$$\omega=\{(x,\ y):x\sin\theta+y\cos\theta=a\}$$

现有一个图象，问在它上面有没有线？这在特征空间中可以解决，在边界上每一点都画出一条直线，在特征空间上交点颜色就较深，例



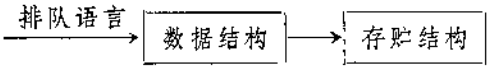
若特征空间是三维的则对应三维空间中的长度不同的线段



只要画出这种图就知道有没有线了。

关于数据结构问题

现在计算机在美国最大的应用不是用来计算，而是用在管理上，最大的问题是数据和信息的存贮和检索。计算主要已是数值分析的问题而不是机器的问题。计算机的主要发展是信息系统方面。一个信息系统可由二步组成



数据结构是机器外的形式，存贮结构是机器里的形式。大的机器象一个“数据银行”，大家都往里存放，但又不能发生冲突，这叫数据安全问题。最近 IBM 公司的总经理讲下五年最大的问题是数据安全问题。大机器的主要问题是软件，小机器的主要问题是硬件。

数据结构的最基本的概念是表征。

例如：年龄取值 {0～100}，于是 A_i=(名子 (A_i), D_i) 是个集合。上式右部可能有个代数顺序，可能有个域或者环。

表征有两种看法：

一种叫关系的模型：

先有一个群的表征 A_1, A_2, …, A_n

每一个表征都有一个状况。

每个取值范围 D_1, D_2, …, D_n

决定了一个乘积空间的子集 $D_1\times D_2\times\cdots\times D_n$。

例：

学生姓名	学生号码	学生年龄	读什么科
A	1	20	工　程
B	2	19	物　理
C	3	20	物　理
X	4	22	工　程
Y	5	18	数　学
Z	6	20	工　程

每一行为一点，每一列为一函数，行的号码即自变量。第二个看法是体系的模型，主要的概念是群。

群＝(群的名字，{群的结果}，{其它的群})

这个定义是递归的。

例：{学生，{学生名字，年龄}{教学大纲})

{大纲。{Year Department}，{Schedule})

{Schedule，{time，(ourse))

这样就形成了一个树型结构，这种树型结构在计算机上是很方便的，所以现在软件大部分是用这种而不是前者。下面举例：

关系模型	体系模型
数据库＝A Collection relation R_1, R_2, …, R_n 数据存放是相重的。	数据库＝A Collection of Groups, possibly C_{01}
好处：简单，没有递归。可以使用很复杂的排队语言。存取控制方便。	好处：紧致，可以快速检索，找起来很方便。

关系模型真正作是太大了，太困难了。最早应用是航空公司定票子。将来发展是要搞关系模型，现在搞还很困难。现在多是体系模型。

基本问题是根据内容来进行记录和检索。有两种办法。

(1) 地址计算：这地址是取决于内容的。

1) 地址取决于内容。建立一个顺序，有一个表征的关系。

2) 散射编码。

(2) 编索引：规定一个地址表。

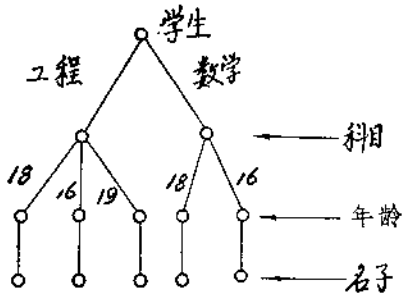
方法 (2) 是有限的，因为造一个地址表就要用存贮，而没有发挥计算机本身可以快速计算的本事。要想快速进行应用方法 (1)。美国有一个软件系统，地址表比数据库大三倍，存贮要求太大了。这是方法(1)的缺陷。

有一个建议：通过一个例子说明。

R = 学生。

学生名字	年龄	科目
A	18	工程
B	18	数学
C	16	工程
D	19	工程
E	16	数学

可在机器上用一个树型结构表示



这样存贮可以直接存取，内容相互嵌套，用不同的信息时可以直接从右边入口进来。所以既相当紧凑，也有一个索引。

座　　谈

问： 图象识别中的统计方法请再介绍一下

答： 统计方法一个困难是没有模型。太半是用距离、矩阵。是和统计有关的，但不是全部统计。每一个字在测度空间中都有一个点。困难是没有一个标准。不是完全是统计，有一半是统计。开始时要有许多假定，分布怎样……结果是趋近的。对统计在理论上是有研究的，但实际上不一定有用。机器是很

粗，在1966年 IBM 公司用 16×16 一方框七个变量，100个布尔函数。大半是局部的。100个选出来花很多时间。

问： 最近实用的机器有没有？

答： IBM 公司文字识别不研究了。

问： 减少差错用实验？

答： 是的。给政府机关用，只做了一架。另一厂做了一台，价格为70万美金。IBM 公司价格贵，实验费贵。

问： 现在手写体的识别怎样？

答： 真正的手写不大可能，银行里的签名是用磁性墨水印出来对应起来的。手写方块的印刷体知道有人在做。但不如打字的好。看医学图如心电图、脑电图比较好，波形比图象容易，波形是一维的，图象是二维的。电视扫描不是一种好的办法。

问： 对电脑有何看法？

答： 这是一个有争议的问题。上月我们学校请了三个人，他们争起来了。一个是 L. S. Coles 认为做头脑只是时间问题。第二是 Weisenbaum 是做软件的，很有地位。也认为可以。但认为对社会讲不应该做。第三是 Dreyfus 属于保守，认为一是不可能，二是不必要。

计算机要很大，把一个人的经历、社会的环境，文明的环境放近去。关键在于脑子怎样做工作的不清楚。

问： 在美翻译用的机器有没有？

答： 有人在研究，无商品。对于翻译的结果都很失望。

做翻译不可能，但帮助人可能。目前有机器字典。

问： 生物学家对电脑看法

答： 懂的人不太多。有一个人找到了一个动物，眼睛十分简单，研究其神经感官。有收获。十年之内很多问题会有发展。生物学最近发展非常非常快。计算机从电子管发展到集成电路，电路发展很快，磁芯没什么变化。有一个公司，3片就是一个计算机。